

経路検索ログデータを用いた突発的検索行動の分析

鳥取大学 学生会員 ○小田島輝知
 鳥取大学 正会員 桑野 将司
 鳥取大学 非会員 森山 卓
 鳥取大学 学生会員 細江 美欧

1. 序論

交通統計の手法は時代とともに改良が加えられてきたが、交通センサスやパーソントリップ調査などは依然として代表値の把握に留まっている。ここでの代表値とは平均的な交通行動を示すものであり、従来の交通統計手法では突発的な交通行動の分析には重点が置かれていない。本研究では「突発的」な状態について、日常的な状態と比べ検索者が多い、あるいは少ない状態と定義する。公共交通において、突発的な状態の発生は交通環境の安全性と合理性を脅かす可能性があり、突発的移動需要の要因を解明することは重要である。

突発的な状態を分析するにあたり、交通系 IC カードのログデータに代表される実行データを用いると、例えば当初利用を希望していた便に満員のため乗れない、といった利用を希望しできなかった人のデータを考慮できないという問題が生じる。この問題に対して、経路検索システムのログデータを用いることで利用できなかった人の行動も捉えることができる。

本研究では状態空間モデルを用いて、経路検索システムの過去の総検索件数から突発的検索行動があった日を抽出する。その後、抽出された日の突発的移動需要の要因について考察を行うことを目的とする。

2. 使用データの概要と基礎集計

A) 使用データの概要

本研究では、神姫バスグループが提供するバス事業者版経路検索システム「神姫バス NAVI」の2018年10月1日から2019年9月30日までの1年間分のログデータを用いる。神姫バスグループの詳細については、ホームページを参照されたい。

神姫バス NAVI には経路検索機能、時刻表検索機能など複数の機能がある。これら機能別の検索割合は、日によって大きな差異はみられなかった。そこで、本研究では、全機能への総アクセス件数（以降、日別システム総

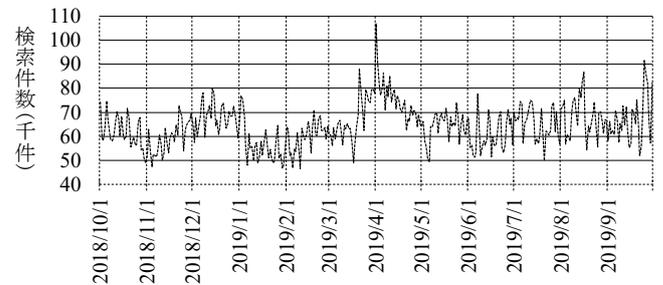


図-1 日別システム総検索件数

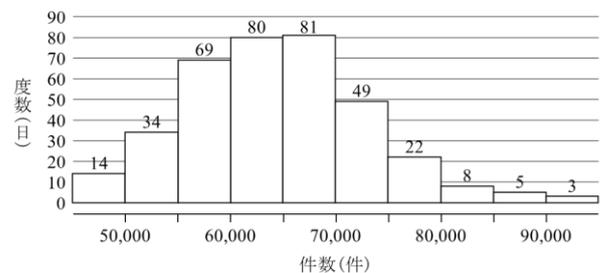


図-2 日別システム検索件数の分布

検索件数)を分析対象とする。

B) 基礎集計

図-1 に1年間の日別システム総検索件数の推移を示す。1年間では延べ23,508,476件の検索があった。1日当たりでは平均約64,407件の検索があり、中央値は2019年5月21日(火)の64,055件、最も検索件数が多い日は、2019年4月1日(月)の107,073件、最も検索件数が小さい日は2019年2月10日(日)の46,575件である。

図-2に日別システム検索件数の分布を示す。1日当たりの検索件数の分布は、65,001件~70,000件の度数が最も大きい。検索件数上位8日にあたる85,001件以上の日は2018年10月1日(月)、2019年の1月2日(水)、3月21日(木・祝)、4月1日(月)、5月6日(月・祝)、2019年の6月7日(金)、9月23日(月・祝)、9月25日(水)である。

キーワード 経路検索システム, ログデータ, 時系列分析

連絡先 〒680-8550 鳥取県鳥取市湖山町南4丁目101番地

TEL 0857-31-5313

3. 突発的検索行動抽出のための統計的手法

本研究では、状態空間モデルを用いて突発的検索行動の抽出を試みる。時系列モデルである状態空間モデルは、データの変動分析に適しているといわれる。

$n = 1, 2, \dots, T$ を時点、 x_n を状態値（ベクトル）、 y_n を観測値としたとき、状態空間モデルの一般式は式(1)の状態方程式と式(2)の観測方程式で表現できる。観測値 y_n に加えて状態値 x_n を仮定することが状態空間モデルの特徴である。

$$x_n = F_n x_{n-1} + G_n v_n \quad (1)$$

$$y_n = H_n x_n + w_n \quad (2)$$

ただし、

$$x_0 \sim N(x_{0|0}, V_{0|0})$$

$$v_n \sim N(0, V_v)$$

$$w_n \sim N(0, \sigma^2)$$

F_n, G_n, H_n : 各時点 n において決まる定数行列

$x_{\cdot|j}$: 条件付き平均 ($x_{j|k} = E[x_j | y_{1:k}]$)

$y_{1:k}$: 時点 $n = k$ までに得られた観測値全体

$$(y_{1:k} = \{y_1, y_2, \dots, y_k\})$$

$V_{\cdot|j}$: 分散共分散行列

$$(V_{j|k} = E[(x_j - x_{j|k})(x_j - x_{j|k})^T])$$

v_n : システムノイズ行列

V_v : システムノイズ行列の分散行列

w_n : 不規則成分

σ^2 : 不規則成分の分散

状態空間モデルは状態値の遷移と観測値と状態値の関係がマルコフ過程に従うことを仮定している。本研究では、すべての状態値と観測値は正規分布に従うと仮定する。カルマンフィルタアルゴリズムにより、各時点の値を求める。次の時点を予測した後にフィルタリングを行い、観測値を用いて状態値の補正を行う。

[次時点の予測]

$$x_{n|n-1} = F_n x_{n-1|n-1} \quad (3)$$

ただし、

$$V_{n|n-1} = F_n V_{n-1|n-1} F_n^T + G_n V_v G_n^T \quad (4)$$

[フィルタリング]

$$x_{n|n} = x_{n|n-1} + K_n (y_n - H_n x_{n|n-1}) \quad (5)$$

ただし、

$$K_n = V_{n|n-1} H_n^T (H_n V_{n|n-1} H_n^T + \sigma^2)^{-1} \quad (6)$$

$$V_{n|n} = (I - K_n H_n) V_{n|n-1} \quad (7)$$

このようにして得られたフィルタ分布は、 n 時点までのすべての観測値を用いた状態の推定となる。

次時点の予測とフィルタリングにより n 時点すべてのパラメータが求まったところで、最終時点から順に時点を戻りながら、改めてパラメータを計算する平滑化を行う。これらの処理を、式(8)から式(10)に示す。

$$x_{n|T} = x_{n|n} + A_n (x_{n+1|T} - x_{n+1|n}) \quad (8)$$

ただし、

$$A_n = V_{n|n} F_n^T V_{n+1|n}^{-1} \quad (9)$$

$$V_{n|T} = V_{n|n} + A_n (V_{n+1|T} - V_{n+1|n}) A_n^T \quad (10)$$

状態空間モデルには多様なバリエーションがあるが、本研究ではローカルモデルとローカル線形トレンドモデルを用いる。ローカルモデルでは、平均を表すレベル成分が常に一定値であると仮定する一方で、ローカル線形トレンドモデルでは、レベル成分に線形トレンドを考慮する。また、レベル成分およびトレンド成分はそれぞれ確定的および確率的なものが存在する。周期要素のあるモデルでは、さらに元の時系列データに周期変動があることを仮定する。周期成分にも確定的または確率的のそれぞれを仮定することができる。本研究では、これらの組合せから考えられる複数のモデルを推計・比較し、赤池情報量基準 AIC によって、最適なモデルを決定する。

以上を定式化すると、本研究で使用する観測方程式は次のように表される。

$$y_n = s_n + d_n + w_n \quad (11)$$

ただし、

s_n : ローカルレベル成分

d_n : 周期変動成分

w_n : 不規則成分

トレンド t_n を含むレベル成分 s_n は長期的な傾向を表現する。

$$s_n = s_{n-1} + t_{n-1} + v_{n1} \quad (12)$$

$$t_n = t_{n-1} + u_n \quad (13)$$

ただし、

$$v_{n1} \sim N(0, \tau_s^2)$$

$$u_n \sim N(0, \tau_t^2)$$

t_n : ローカルトレンド成分

v_{n1} : ローカルレベル成分の誤差成分

u_n : ローカルトレンド成分の誤差成分

τ_s^2 : ローカルレベル成分の誤差成分の分散

τ_t^2 : ローカルトレンド成分の誤差成分の分散

周期変動成分 d_n は季節や曜日のような繰り返し現れる変動成分を表現する。事前分析として、1年間の検索

件数の推移に対し自己相関係数を算出した結果、ラグが7のときに高い正の相関が確認された。使用データ期間が1年であることから、月変動や季節変動は考慮できないため、本研究では周期変動は曜日による影響のみを想定し、以下の通り、周期を7日とする。

$$\sum_{i=0}^6 d_{n-i} = v_{n2} \quad (14)$$

ただし、

$$v_{n2} \sim N(0, \tau_d^2)$$

v_{n2} : 周期成分の誤差成分

τ_d^2 : 周期成分の誤差成分の分散

このとき、各成分で構成される状態ベクトル $x_n = (s_n, t_n, d_n, d_{n-1}, \dots, d_{n-5})^T$ には一期前のみに基づく状態方程式が成り立つ。

次に得られる各モデルのAICを求める。各モデルのパラメータとAICの結果を表-1に示す。なお、表-1において、レベル成分及び周期成分が確定的な場合はそれぞれ“ $v_{n1} \equiv 0$ ”、“ $u_n \equiv 0$ ”と表現している。

分析の結果、「C-2 確定的周期要素を持つローカル線形トレンドモデル」のAICが最小となった。このモデルを用いて得られた不規則変動から、突発的検索行動について分析する。次章ではこの不規則変動から突発的状態が発生した日を抽出し、さらにその要因について考察を行う。

4. 突発的検索行動の抽出と要因の考察

A) 状態空間モデルによる突発的検索行動の抽出

図-3の対数変換した日別システム総検索件数に対して、状態空間モデルを適用した結果得られたレベル成分、周期成分、不規則成分を図-4、図-5、図-6に示す。図-4のレベル成分を見ると、2019年7月頃から多少の増加傾向は見られるものの、期間全体を通じて大きな変動はみられない。これは、長期的な視点からは経路検索システムの検索者数は大きく増減していないことを意味する。レベル成分を期間に分けて考察すると、冬休

表-1 検討したモデルのパラメータとAIC

モデル	レベルの確率成分	傾き成分	傾きの確率成分	周期成分	周期の確率成分	AIC	
A-1	0	0	0	0	0	-414.98	
A-2	v_{n1}					-582.65	
B-1	0					-412.32	
B-2	v_{n1}	t_n	0	0	0	-567.80	
B-3						u_n	-565.80
C-1	0	t_n	u_n	d_n	0	-430.01	
C-2	v_{n1}					-697.58	
C-3	v_{n1}					v_{n2}	-695.58
D-1	0	t_n	0	d_n	0	-420.84	
D-2						v_{n1}	-418.84
D-3						v_{n1}	u_n

みや年度始めは検索者が多く、大きなイベントが少ない平常時は検索者が少ない傾向が明らかとなった。

また、2019年9月末頃に増加傾向にあるが、2019年10月からの増税による運賃値上げがあったため、定期運賃などの関連する情報を収集する検索者数が増えたと考えられる。

図-5に示す周期成分には、確率的要素が含まれていないため、年間を通じて曜日ごとに一定の値を持つ。周期成分の曜日ごとの成分値を示した結果を図-7に示す。図-7から、月曜日から木曜日までは目立った特徴は無い一方で、金曜日から日曜日にかけて周期成分は特徴的な変動を示している。金曜日次いで土曜日が正で大きな値を示し、日曜日は負の値を返している。古川ら⁹⁾は、経路検索システムの利用者は、公共交通を利用する直前に検索する人もいれば、余裕をもって事前に検索する人もいることを指摘しており、特に観光目的と考えられる検索は、数日前に検索される傾向があることを明らかにしている。古川らの考察を参考に、周期成分には、行楽の前日に検索する休日利用者と、ビジネスを理由に直前に検索する平日利用者の検索時期の違いが現れていると考えられる。

以上、日別システム総検索者数の変動から、レベル成分と周期成分で説明できない不規則成分(図-6)に突発的検索行動が含まれていると考え、次節で考察する。

B) 突発的検索行動の要因の考察

図-6の不規則成分から、突発的検索行動が発生したとみられる日を抽出したものを図-8に示す。突発的と判断する基準は、不規則変動の絶対値が0.1以上をとるものとした。この結果、突発的検索行動が発生した日は分析対象の1年間に18日あった。そして当該日およびその前後数日について、天候、催し物、学校行事などを中心にウェブや過去の新聞記事から情報を集めた。この調査の結果、突発的検索行動が発生した18日のうち11日について要因を推察することができた。

2019年9月23日(月)、25日(水)、29日(日)、30日(月)に大きな突発的変動を示している。同26日(木)および30日(月)には神戸市内でラグビーワールドカップが開催されており、検索件数の突発的な増加に影響したと考えられる。

2018年12月31日(月)、2019年1月2日(水)、4月1日(月)も突発的検索行動が発生した日として検出されている。一般に大晦日の12月31日と元旦の翌日である1月2日は

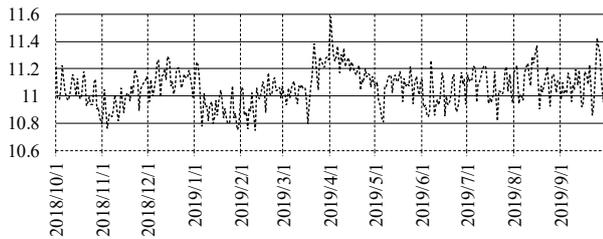


図-3 対数変換した日別システム総検索者数

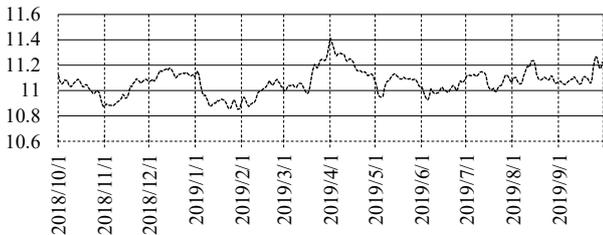


図-4 レベル成分

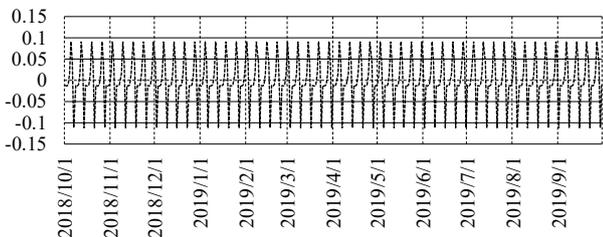


図-5 周期成分

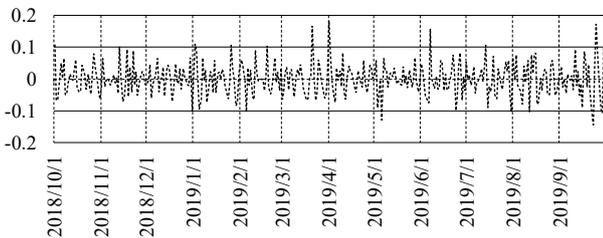


図-6 不規則成分

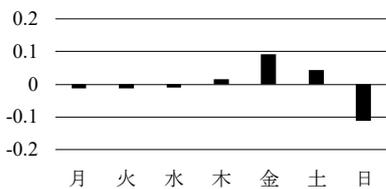


図-7 曜日ごとの成分

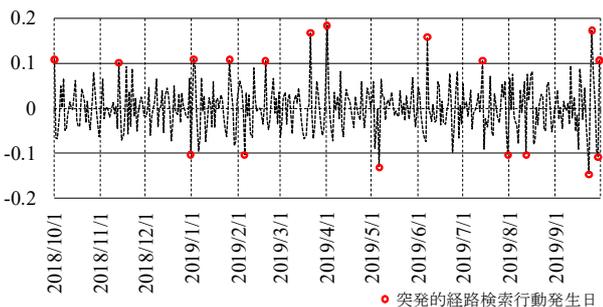


図-8 突発的検索行動の抽出結果

道路交通において交通渋滞の特異日と認知されており、また、4月1日(月)は年度始めであることから、それぞれ不規則成分が大きくなったと考えられる。

次に、突発的変動の最も多い要因として考えられる天候に着目する。調査の結果、大型台風が通過する前後、降雪、および激しい降雨があった日に突発的成分が大きいたことが明らかとなった。これにあてはまるのは、台風が通過した2018年10月1日(月)、2019年7月27日(水)、2019年8月14日(水)の前後、西日本広く降雪があった2019年1月26日(土)、近隣の広島県などで避難勧告が発令されるほど激しい降雨があった2019年6月7日(金)である。ここで、天候が悪化した際に検索件数が増える理由は大きく2つ考えられ、1つ目は、公共交通機関のダイヤの乱れを予想し平常時と異なる時刻での移動を望むため、2つ目は、神姫バスNAVIにある現在のバス位置を調べる機能の需要が高まるためと考えられる。

5. 結論

本研究では、神姫バスグループが提供する経路検索システム「神姫バスNAVI」のログデータから突発的な検索行動があった日を抽出し、その検索行動の要因を考察した。本研究では状態空間モデルを用いた不規則変動成分の抽出を行った。不規則成分から抽出された突発的検索行動の発生要因を、ウェブと過去の新聞記事を用いて考察した。状態空間モデルの適用により、突発的検索行動は対象の1年間のうち18日抽出された。この18日について要因の調査を行ったところ、総検索件数は広範囲に影響を及ぼす天候の悪化と大規模な催し物から影響を受けていることがわかった。

本研究では突発的検索行動が抽出された18日のうち7日において明確な要因を特定できなかった。これらの課題に対する解決策として、多様なバリエーションを組むことができる状態空間モデルをさらに拡張し、新たな変動要因を考慮すること、本研究で考慮しなかった起終点情報を考慮することが挙げられる。これらの解決策によって、突発的移動需要の要因解明につながる可能性がある。

参考文献

- 1) 神姫バス株式会社：ホームページ、<https://www.shinkibus.co.jp/>
- 2) 萩原淳一郎、瓜生真也、牧山幸史：基礎からわかる時系列分析—Rで実践するカルマンフィルタ・MCMC・粒子フィルタ、技術評論社、2018。
- 3) 中村 和幸：統計の時系列解析2：状態空間モデル(チュートリアル)、応用数理、Vol. 24, No. 2, pp. 74-79, 2014。
- 4) 古川ゆり、桑野将司、秋元美穂奈、菅原一孔：経路検索履歴データを用いた事前検索時間の分析、交通工学論文集、Vol.5, No.2, pp. A_184-A_192, 2019。